Re-Representation and Creative Analogy:

A Lexico-Semantic Perspective

Tony Veale,

School of Computer Science and Informatics,

University College Dublin,

Ireland.

Tony.Veale@UCD.ie

Abstract

Analogy is a powerful boundary-transcending process that exploits a conceptual system's ability to perform controlled generalization in one domain and re-specialization into another. The result of this semantic leap is the transference of meaning from one concept to another from which metaphor derives its name (literally: *to carry over*). Such generalization and re-specialization can be achieved using a variety of re-representation techniques, most notably abstraction via a taxonomic backbone, or selective projection via structure-mapping over propositional content. In this paper we explore both the extent to which a bilingual lexical ontology for English and Chinese, called HowNet, can support each technique, and the extent to which both are, ultimately, variations of the same process of creative re-representation.

Keywords: Analogy, Abstraction, Structure-Mapping, Ontologies, Lexical Semantics

1 Introduction

Given a recalcitrant category that unduly limits one's actions, the creative individual seeks a new category within which to operate. Indeed, this ability to reconceptualize an object or idea from a different perspective, or from within the viewpoint of a different category, is conventionally considered central to human creativity. Analogy is just one manifestation of creative cognition, but one that clearly illustrates this ability of creative thinkers to transcend conceptual boundaries and perform a semantic leap from one category structure to another (e.g., see Veale, 2003,2004b). Reconceptualization can be dramatic, as when a scientist is forced to accept a paradigm shift from one theory to another (see Kuhn, 1962), or mundane, as when someone uses a credit-card to open a door, a screw-driver to open a can of paint, or a chair to wedge a door shut. Humour also employs reconceptualization as a resolution mechanism: consider how many jokes stretch or transcend the definitional boundaries of conventional categories[1], or employ a punch-line that forces a listener to recategorize his or her interpretation of the preceding narrative (e.g., see Ritchie, 1999; Attardo et al., 2002; Veale, 2004a).

Reconceptualization is certainly a good high-level *story* of what occurs in creative situations, but as computationalists, we require a more specific account. In particular, we require an algorithmic insight into what it means to transcend category boundaries, and this in turn requires some minimal commitment to some form of conceptual

[1] For instance, many jokes play with the boundaries of taboo categories to categorize non-taboo events – like visiting the doctor – in terms of taboo events such as a sexual infidelity (see Attardo *et al.*, 2002). Alternately, some jokes re-imagine socially mediated categories, like those that constitute our value systems, in subjective terms. Consider the following remark from a famously talented, and famously dissolute footballer: "I spent most of my money on alcohol, women and gambling, and the rest I wasted."

representation. In fact, computationalists of an empirical bent often prefer to work from the latter to the former: given a particular resource with a specific conceptual representation, empiricists attempt to frame the problem in terms of this representation. Perhaps unsurprisingly, this is precisely what we shall attempt to do in this paper. The resource we focus on here is HowNet, a bilingual lexical ontology for Chinese and English (see Dong, 1988; Wong, 2004). Earlier experiments (e.g., see Veale, 2004b, 2005) suggest that HowNet is well suited to the demands of analogy reasoning, and thus some forms of creative reasoning, since HowNet combines a taxonomic backbone with an explicit, if somewhat sparse and under-specified, propositional semantics.

Now, theories of analogy and metaphor are typically based either on structure-mapping (e.g., see Falkenhainer *et al.*1989; Veale and Keane, 1997) or on abstraction (e.g., see Hutton, 1982; Fass, 1988; Way, 1991; Veale, 2003). While the former is most associated with analogy, the latter has been a near-constant in the computational treatment of metaphor. Structure-mapping assumes that the causal behaviour of a concept is expressed in an explicit, graph-theoretic form so that unifying sub-graph isomorphisms can be found between different propositional representations. In contrast, abstraction theories assume that analogous concepts, even if far removed in ontological terms, will nonetheless share a common hypernym that will capture their causal similarity. Thus, we should expect an analogous pairing like *cancer* and *assassin* to have very different immediate hypernyms but to ultimately share a behavioural abstraction like *kill-agent* (e.g., see Veale, 2003).

With a well known lexical ontology like WordNet (see Miller, 1995), both structure-mapping and abstraction-based approaches are problematic. The idea that a

one-size-fits-all representation like WordNet will actually provide a hypernym like *kill-agent* seems convenient almost to the point of incredulity. As much as we want our ontologies to anticipate future analogies with these pro-active categorizations, most off-the-shelf ontologies simply do not possess such convenient terms (see Wong, 2004). Similarly, WordNet lacks the propositional content that is the necessary grist for a structure-mapping approach. The semantic content that would ideally fill this role is not explicit, but implicitly resides in the unstructured textual glosses that annotate each lexical concept.

In this paper we explore the extent to which another lexical ontology, the aforementioned Chinese/English HowNet system (see Dong, 1988; Carpuat *et al.* 2002; Wong, 2004) supports the kind of reconceptualization that is required in the generation and interpretation of creative analogies. The WordNet-like taxonomic backbone, in combination with its own unique propositional semantics, allows us to evaluate the extent to which both structure-mapping and abstraction theories of analogy can be supported by the same lexical ontology.

We begin by briefly summarizing past approaches to the computational treatment of metaphor and analysis in section 2, before comparing the pros and cons of WordNet and HowNet in section 3. In section 4 we describe a form of reconceptualization that relies on conceptual abstraction; however, we do not propose a model of simple taxonomic abstraction, but one of *relational* abstraction, since only the latter allows us to generalize over the functional and behavioural meaning of a concept. To extend the reach of relational abstraction to representations that would not otherwise support this technique, we also present here a form of representational transformation called

structural inversion. This is, in essence, a form of figure-ground reversal in which alternative representations for an under-specified concept can be sought by turning to elements in the conceptual background. In section 5 we then describe a form of reconceptualization based on structural rarefaction; this in turn supports a structure-mapping approach to analogy using HowNet representations. In section 6, both of these approaches to reconceptualization are subjected to a comparative evaluation across the entirety of HowNet. We conclude by arguing, on the basis of this evaluation, that these approaches are ultimately complementary, inasmuch as a synthesis of both produces better performance than does either approach in isolation.

2 Past Work

That analogy and metaphor operate across multiple levels of conceptual abstraction has been well known since classical times. Aristotle first provided a compelling taxonomic account of both in his Poetics (see Hutton, 1982 for a translation), and computationalists have been fascinated by this perspective ever since. While the core idea has survived relatively unchanged, one must discriminate theories that apparently presume a static type-hierarchy to be sufficient for all abstraction purposes (e.g., Fass, 1998), from theories that posit the need for a dynamic type hierarchy (e.g., Way, 1991; Veale, 2003). One must also differentiate theories that have actually been implemented (e.g., Fass, 1988; Veale, 2003,2004) from those that are either notional or that seem to court computational intractability (e.g., Hutton, 1982; Way, 1991). Perhaps most meaningfully, one must differentiate theories and implementations that assume

hand-crafted, purpose-built ontologies (e.g., Fass, 1988) from those that exploit an existing large-scale resource like WordNet (e.g., Veale, 2003,2004). The latter approach side-steps any possible charge of hand-crafting by working only with third-party resources, but at the cost of living with their perceived flaws and inadequacies.

Structure-Mapping theory is founded on the premise that the most satisfying analogies are those that operate at the causal level of representation, since causality allows an analogy to offer a deep explanation for a poorly understood phenomenon (e.g., see Falkenhainer *et al.* 1989). Thus, *the atom as miniature solar-system* is a satisfying analogy because both source and target are causally structured around the notion of rotation. Furthermore, when comparing agents or artefacts (e.g., see Veale and Keane, 1997), this causality can be captured by considering the functional or behavioural commonality between target and source: a footballer can be meaningfully described as a gladiator or a warrior since each exhibits competitive behaviour, and a scalpel can be compared to a sabre, a sword or a cleaver since each has a cutting behaviour.

By employing a single lexical resource, HowNet, to implement both the relational abstraction *and* the structure-mapping theories of analogy, we have as a secondary goal a demonstration that both perspectives are not fundamentally opposed. Structure-mapping can be seen as a form of structural-abstraction, where one abstracts out the causal backbone of a concept, while taxonomic abstraction, if performed upon the relations implied by a concept rather than the concept itself, can also be seen as a highly selective form of structure-mapping. Ultimately, both kinds of approach attempt to capture the functional or behavioural commonality between a pair of source and

target concepts: a surgeon can be meaningfully described as a repairman since both occupations have the function of restoring an object to an earlier and better state; the distinction, which is glossed over both by abstraction and structure-mapping approaches, is that a surgeon restores by healing, while a repairman restores by mending.

3 Comparing WordNet and HowNet

HowNet and WordNet each reflect a different view of semantic organization. WordNet is *differential* in nature: rather than attempting to express the meaning of a word explicitly, WordNet instead differentiates words with different meanings by placing them in different synonym sets, and further differentiates these synsets from one another by assigning them to different positions in its taxonomy. In contrast, HowNet is *constructive* in nature. It does not provide a human-oriented textual gloss for each lexical concept, but instead combines sememes from a less discriminating taxonomy to compose a semantic representation of meaning for each word sense.

For example, the lexical concept *surgeon*|医生 is given the following semantic definition in HowNet:

surgeon|医生≡*{human*|人*:HostOf={Occupation*|职位*},*

domain={medical|医*}},*

{doctor|医治*:agent={~}}}*

which can be glossed thus: “a surgeon is a human with an occupation in the medical domain who acts as the agent of a doctoring activity.” The *{~}* construct serves as a

self-reference, to mark the location of the concept being defined in the given semantic structure. The oblique reference offered by the tilde serves to make the definition more generic, so that many different concepts can conceivably employ the same definition. Thus, HowNet uses the above definition not only for surgeon, but for medical workers in general, from orderlies to nurses to internists and neurologists.

Perhaps because HowNet relies less on hierarchical differentiation, it has a considerably less developed middle ontology than WordNet. For instance, most kinds of person in HowNet[2], from mathematicians to hobos, are placed directly under the hypernym *human|人*, eschewing the intermediate concepts like *{professional}, {specialist}* and *{worker}* that give substance to WordNet's middle ontology. We note that HowNet does indeed define these concepts – but unlike WordNet, it does so at the leaf level where they add nothing to the internal structure of the taxonomy.

3.1 Analogical Signatures and HowNet

Nonetheless, the skeletal nature of HowNet semantic definitions, combined with the wide-spread use of *{~}* as a generic reference, suggests how HowNet might support an efficient approach to analogical reasoning. By indexing each concept on a reduced form of its semantic definition – an *analogical signature* – analogies will correspond to collisions between concepts with different definitions but with identical signatures. Such an approach can be efficiently implemented using simple string hashing of signatures, to

[2] We note in passing that the Chinese origins of HowNet explains some additional, cultural distinctions between Princeton WordNet and HowNet. For instance, WordNet defines dogs as a kind of canine; HowNet defines dogs as kinds of livestock.

detect analogical collisions between kitchens and factories, generals and admirals, ballet dancers and acrobats, or cruise missiles and arrows. The devil here is in the *lack* of detail: because HowNet's definitions are frequently imprecise and fail to fully specify a concept, they allows others – potential analogues – to occupy the same reduced semantic space. The further we exacerbate this deficiency, indexing each definition on an increasingly diluted version of itself, the more distant and creative will be the analogies that are generated. For example, excluding the hypernym of a definition, or its domain markings, facilitates analogies between people and non-people, such as pests and persecutors, or hackers and viruses.

To implement both the abstraction and structure-mapping theories of analogy, we will explore the effectiveness of two kinds of analogical signatures in this paper: relational signatures derived, via abstraction, from the predicate and case-role of a proposition, and structured, template-like signatures based on generalized propositional content in which place-holder variables may be added.

4 Re-Representation via Abstraction Signatures

Given the general impoverishment of HowNet's middle ontology (at least compared with that of WordNet), abstraction-based signatures should not be based directly on taxonomic organization. Rather, by instead deriving analogical signatures from the relational structure of a concept's semantic definition, we can better capture the functional and behavioral nature of the concepts concerned. We can do this by focusing on how each concept is situated with respect to its relational context, which is to say, by

targeting the explicitly self-referential *{~}* in each definition. For instance, consider the following semantic definition of repairman in HowNet:

repairman|修理工 ≡{human|人:

HostOf={Occupation|职位},

{repair|修理:agent={~}}}

Noting the relational position of *{~}* here, we can infer that a repairman is the agent of a repairing activity. Expressing this as a new taxonomic type, we can reify the combination of activity and role to create a new taxonomic term *repair-agent*, of which repairman will be an instance. From an analogical perspective, *repair-agent* thus serves as a good relational signature for *repairman|修理*.

Further noting that the HowNet taxonomy defines the predicate *repair|修理* as a specialization of the reinstatement predicate *resume*|恢复, we can further establish *repair-agent* as a specialization of *resume-agent*[3]. This double layer of abstraction effectively establishes a new, parallel taxonomy that organizes lexical-concepts according to their analogical potential, rather than their formal taxonomic properties. For instance, as shown in Figure 1, *resume-agent* will encompass not only *repair-agent*, but *doctor-agent*, since HowNet also defines the predicate *doctor|医治* as a specialization of *resume*|恢复.

[3] HowNet uses the predicate *resume* in the sense of *restore*, that is, "to resume an earlier, better state".

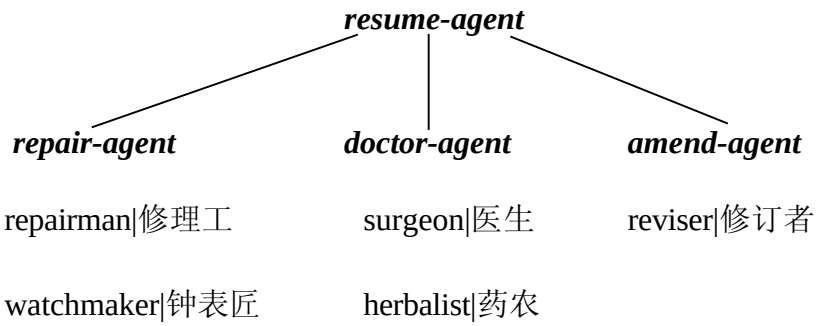

Figure 1: Portion of a new three-level abstraction hierarchy derived from HowNet's relational structures.

Relational signatures like *repair-agent* and *doctor-agent* are, in essence, new lexical concepts that allow particular problems of an analogical or metaphoric nature to be solved creatively. Hierarchies like that of Figure 1 thus reflect the general philosophy of creativity espoused in McCarthy (1999), which stipulates that a solution is only truly "creative" when it recruits or creates concepts that were not directly mentioned in the original problem specification. McCarthy's viewpoint is interesting because it forces us to evaluate creativity not just on the utility of the end-product, which might equally be produced by the most banal of exhaustive searches, but on the selective means through which this end was achieved.

In general, relational signatures are generated as follows: given a semantic fragment *F:role={~}* in a HowNet definition of a concept C, we create the signatures *F-role* and *F'-role*, where *F'* is the immediate HowNet hypernym of *F*, which in turn is the immediate hypernym of *C*. The role in question might be *agent*, *patient*, *instrument*, or any other role supported by HowNet, such as *target*, *content*, etc.

Each concept is thus assigned two different relational signatures: a direct signature

(*F-role*) based on the specific relational structure of the concept, and another more abstract signature (*F'-role*) that is generalized, via taxonomic abstraction, from this direct signature. These signatures effectively form an alternate taxonomy by which the lexical concepts in HowNet can be organized for analogical purposes. Figure 2 illustrates a partial hierarchy derived from HowNet definitions of form-altering tools:

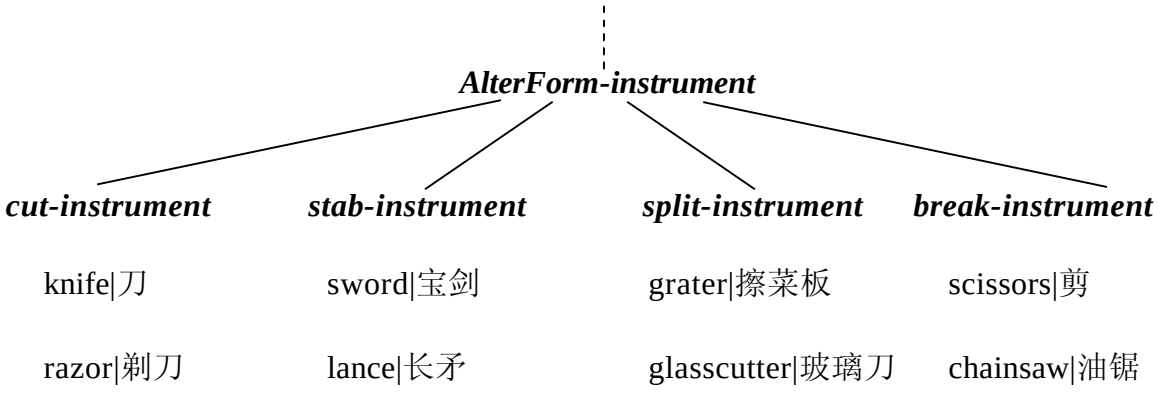

Figure 2: a derived taxonomy of relational signatures that facilitates analogy between instruments that "alter the form" of other objects.

This additional layer of abstraction is necessary to facilitate creative analogy between semantically distant concepts. Nonetheless, we note that since HowNet's designers have already exercised a certain degree of metaphoric license, even concepts with the same direct signature can exhibit a surprising degree of semantic variety.

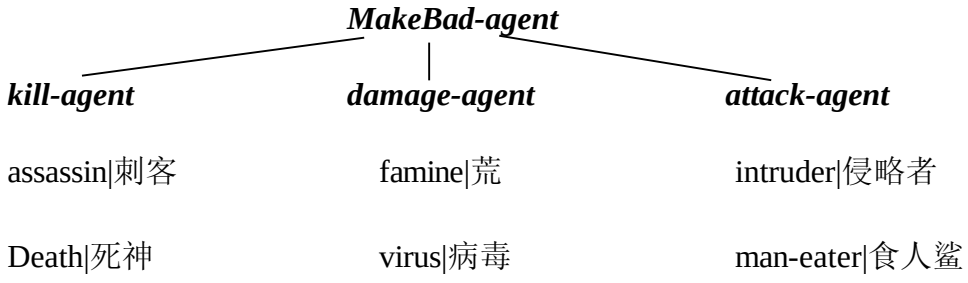

Figure 3: semantic diversity among concepts with the same relational signatures.

This diversity, as illustrated by Figure 3, means that the analogy “Death is an assassin” can be generated in a single generalization step, while the analogy “Death is a man-eater” can be generated with just two generalization steps.

4.2 Reconceptualization via Structural Inversion

Since the partial taxonomies of Figures 1, 2 and 3 do not exist in HowNet, but are derived from HowNet representations, it seems quite meaningful to refer to these taxonomies as reconceptualizations of the original HowNet taxonomy. However, the technique of relational abstraction is seriously limited, in a way we shall explicitly quantify later, by its ability to apply only to those definitions that are self-referential. If no relational signature can be generated for a given concept, as is the case when a definition is not structured around the use of {~}, then no analogies can be retrieved for that concept. For example, consider the HowNet definition of “*bicycle*”:

bicycle|*单车* ≡ {LandVehicle|车: modifier={manual|非自动}}

Clearly, no analogical signature can be derived from this overly under-specified definition. Nevertheless, consider another HowNet entry that refers to this *bicycle*|*单车* definition:

cyclist|*车手* ≡

{human|人:

{drive|驾驭:

agent ={~},

patient={LandVehicle|车: modifier={manual|非自动}}}}

The concept *cyclist*|车手 is clearly much better situated with respect to analogical reasoning, giving rise to the signatures *drive-agent* and its abstraction *CauseToMoveInManner-agent* that are shared by pilot, chauffeur, astronaut and trucker amongst others. Since bicycle occupies the conceptual background of this definition, a figure-ground reversal is needed to bring it into the foreground as the focus of the definition. That is, we can structurally invert this definition to yield an alternate conceptualisation of *bicycle*|单车, by simply replacing the {~} marker with the foreground concept *cyclist*|车手 for which it stands in the definition, thus backgrounding this concept, and replacing the sub-definition of *bicycle*|单车 with {~}, thus foregrounding this concept. This figure-ground reversal is graphically illustrated in Figure 4:

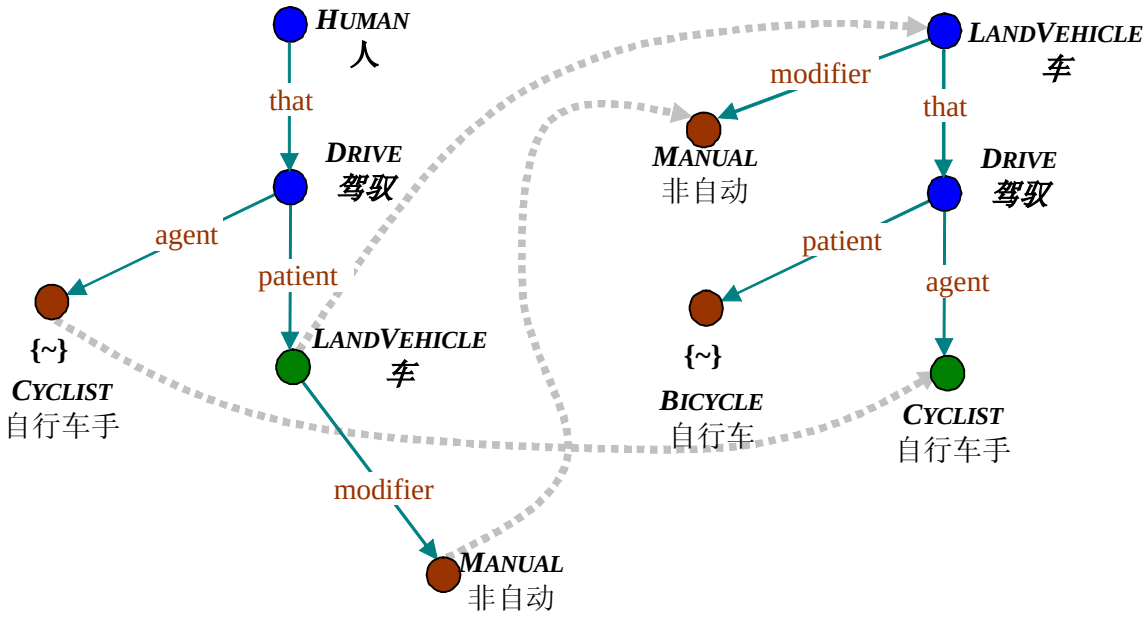

Figure 4: The HowNet definition of *cyclist*|车手 is structurally inverted to yield a richer representation of *bicycle*|单车 than that which is offered by HowNet itself.

Further replacing the taxonomic head of the definition (human|人) yields this new structure:

bicycle|*单车* ≡

{LandVehicle|车:

{drive|驾驭:

agent ={*cyclist*|*车手*},

patient={~}}}

In this reconceptualization, a bicycle is a vehicle that is driven by a cyclist. While this is not a particularly tight definition, it is precisely this lack of formal rigidity that serves to enable creative thinking. After reconceptualization, the concept *bicycle*|*单车* is thus assigned the analogical signature *drive-patient*, facilitating an analogical mapping to boats, airplanes, trucks and even elevators (since HowNet construes each as the patients of a driving activity).

5 Re-Representation via Structural Signatures

The structure-mapping approach also strives for abstraction, not through the selective creation of new taxonyms but through a form of structural *rarefaction*. Recall that structure-mapping theory places particular emphasis on the causal backbone of a concept's propositional content, which is usually projected unchanged from one domain to another (see Falkenhainer *et al.* 1989). Based on this isomorphic alignment of relational structures, the entities contained in each structure are typically placed into a 1-to-1 correspondence with one another. The attributive modifiers of these entities play a more peripheral role in structure-mapping, but in approaches like *Sapper* (Veale and

Keane, 1997) they often serve as a literal grounding for an analogy. Figure 5 depicts an example of the structure-mapping process applied to HowNet representations.

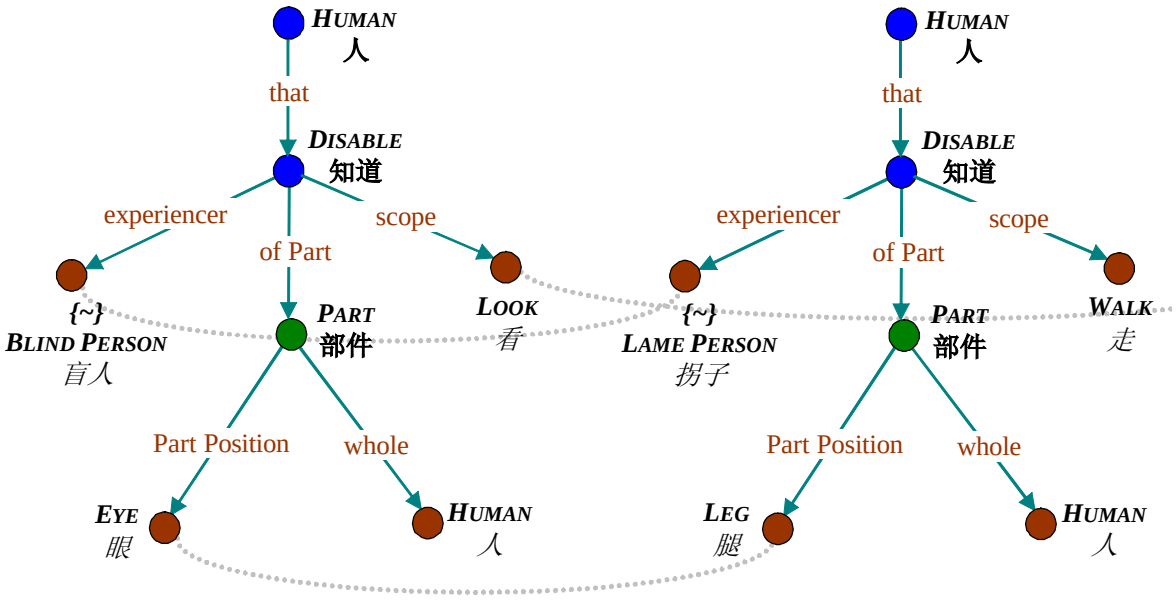

Figure 5: Structure-Mapping applied to the HowNet representations of *blind person*|*盲* and *lame person*|*拐子*, identifying the 1-to-1 entity mappings *look*:*walk* and *eye*:*leg*.

In many cases, the semantic definitions provided by HowNet are already so skeletal and under-specified that we may assume that any structure-mapping signature will preserve the general form or shape of the proposition from which it is derived, one signature per proposition. As an example, consider the HowNet definition of *blind person*|*盲人*:

{human|人: {disable|知道:

OfPart={part|部件:PartPosition={eye|眼}, whole= {human|人}}

experiencer={~},

scope={look|看}}}

In other words, a blind person has “a disability of the eye that affects one’s ability to look”. One finds precisely the same propositional structure in the HowNet definition of *lame person|拐子*, except that *eye|眼* is replaced with l*eg|腿* and *look|看* is replaced with *walk|走*. The goal of a structure-mapping approach is to capture precisely this semantic isomorphism while simultaneously identifying entity-level differences like *eye:leg* and *look:walk* as cross-domain counterparts. We thus need to generalize from each proposition in every definition a structural signature that, by virtue of being identical to another, signals a structural equivalence between the underlying definitions. For instance, the shared signature for *blind person|盲人* and *lame person|拐子* looks like:

{?:{ill|病态:OfPart={?},experiencer={~},scope={?}}}

Generalized structural signatures of this form are generated via a 7-step process:

1. Split each definition into multiple propositions, and generate a separate signature for each.
2. If a proposition describes a noun concept, replace its taxonomic head with a *?* marker (E.g., human|人 → ?). In contrast, if a proposition describes a verb concept, replace its taxonomic head with its most specific hypernym (e.g., *repair* → *resume*).
3. Replace the conceptual arguments bound to each case-role of a predicate with the variable marker *{?}*. These markers will indicate positions in the signature where 1-to-1 correspondences between source and target structures can be made.
4. When a propositional sub-structure corresponds to the definition of another concept, replace the entire sub-structure with a *{?}* variable marker as in 3 above.

5. Replace predicates by their immediate hypernyms in the HowNet taxonomy. Thus, both *repair*|*修理* in the definition of *repairman*|*修理*, and *doctor*|*医治* in the definition of surgeon|医生, should be replaced by the hypernym *resume*|恢复 when generating their respective signatures.

6. Remove any explicit domain tag in a proposition from the corresponding signature (e.g., the assignment *domain={medical*|*医}* in the definition of *surgeon*|*医生*). This is necessary since analogy is meant to transcend domain boundaries.

7. Generalize the value of any purely attributive relation, like *modifier, manner, restrictive, host* or *content*, to its immediate hypernym, and ensure that step 3 above does not variablize the resulting value but allows it to remain present as a literal.

Not all of these steps need to applied to produce a valid signature. For instance, multiple signatures at different levels of detail can be generated for the same proposition by alternately applying or ignoring steps 4 and 5. Indeed, because a different signature is generated for each sub-proposition (except for empty propositions, as we shall discuss below), a given HowNet definition will often generate several structural signatures, so that overall, there may be more unique signatures than unique propositional structures.

Following these 7 steps then, the following structural signatures will be assigned to each of the concepts *surgeon*|*医生, repairman*|*修理, reviser*|*修订者, watchmaker*|*钟表匠* and *herbalist*|*药农*:

{?:HostOf={?}} and *{?: resume*|恢复*:agent={~}}*

However, because the *HostOf* relation always occurs with the binding *Occupation|职位* in HowNet, it is effectively useless as an analogical index and the resulting signature is discarded. So in the example above, only the latter signature is retained.

More structural richness is exhibited by the lexical concepts *apostle|使徒* and *insider|局内人*, whose HowNet definitions are shown below.

apostle|使徒

≡ {human|人:

{believe|修理:

agent={~},

content={humanized|拟人},

domain={religion|宗教}}}

person who knows inside story|个中人

≡ {human|人:

{know|知道:

agent={~},

content={fact|事情:

modifier=covert|隐秘}}}

These are also assigned the same structural signature:

{?: {HaveKnowledge|有知:agent={~}, content={?}}

The sub-structure *{fact|事情:modifier=covert|隐秘}* is completely variablized within the signature of *person who knows inside story|个中人* since this corresponds to the HowNet definition of *secret|秘事* (see step 4). Analogically then, an apostle is a religious insider, one who knows the inside scoop on a given deity (denoted *humanized|拟人* in HowNet).

6 Comparative Evaluation

Consider first the composition of the HowNet version used in this research. It contains 95,407 unique lexical concepts (excluding synonyms) and 23,507 unique semantic definitions. Clearly then, these definitions are under-specified to the extent that many are shared by non-identical concepts (such as *cart|板车* and *bicycle|单车*, which HowNet simply defines as manual vehicles with the same under-specified definition). Furthermore, 90% of these definitions comprise a single proposition, while only 8% comprise two propositions and a mere 2% comprise three or more propositions.

We evaluate the abstraction and structure-mapping approaches using four criteria: *coverage* – the percentage of unique HowNet definitions from which a valid signature can be derived; *recall* – the percentage of unique definitions (not concepts) for which at least one analogical counterpart can be found; *parsimony*– the percentage of effective signatures that can actually be used to generate analogies (the most parsimonious approach is precise in generating only those signatures that are analogically useful); and

richness – the complexity of the mappings captured by each analogy, as measured by the average number of entity correspondences per analogy.

6.1 Evaluating Relational Abstraction

6.1.1 Abstraction Coverage

Since relational signatures exploit occurrences of *{~}* for their generation, both the coverage and recall of the relational abstraction approach depend crucially on the wide-spread usage of this reflexive construct.

However, of the 23,507 unique definitions in HowNet, just 6430 employ this form of self-reference. The coverage offered by relational signatures is therefore just 27% of the available definitions. However, structural inversion enlarges the HowNet semantic space from 23,507 unique definitions to 24,514, with each of these additional 1007 definitions employing {~} self-reference. The coverage of analogical mapping with structural inversion is thus 31% (which represents a 15% improvement).

We note that while 31% is still rather low, the use of {~} is not uniform across HowNet's definitions. The most useful concepts from an analogical perspective, Person, Animal and Artefact, are more densely represented by self-referential definitions than the ontology as a whole, offering 65%, 68% and 42% coverage respectively.

6.1.2 Abstraction Recall

From those definitions containing a {~} self-reference, 1579 unique direct signatures are generated. In turn, another 838 abstracted relational signatures are generalized from

these using HowNet's taxonomic organization of verbs. In total, 2219 unique relational signatures are generated. This reveals that the sets of direct and abstracted signatures are not disjoint, and that in 8% of cases, the abstracted signature of one definition corresponds to the direct signature of another.

The overall recall rate is 30% (or 26% without structural inversion), which is to say, a relational signature enables the recall of at least one analogous definition for 30% of the unique definitions in HowNet. The most productive relational signature is *control_agent*, which serves to analogically co-index 210 unique HowNet definitions, among them the definitions of Boss, Manager, Manipulator, Bosun and Traffic-Cop.

6.1.3 Abstraction Parsimony/Precision

Overall, 1,315 of all 2219 relational signatures prove to be useful in co-indexing two or more definitions, while 904 relational signatures are associated with just a single definition. The parsimony of the abstraction approach is thus 59%, which is to say that 59% of the generated signatures are analogically useful, while 41% serve no analogical purpose and are ultimately rejected. This measure of parsimony is a useful index of predicate re-use in HowNet: a high parsimony score suggests that most definitions are defined using a communal set of predicates that systematically apply to more than one concept; a low parsimony score suggests that most definitions are defined on an ad-hoc basis. A parsimony score of 59% is moderate, suggesting strong systematicity but some ad-hoc tendencies in HowNet.

6.1.4 Abstraction Richness

Since the abstraction approach produces atomic, rather than structured signatures, it is capable of generating only one mapping per analogy, at the gross level of the source and target concepts themselves. For instance, while the abstraction approach can recognize that *blind person|盲人* and *lame person|拐子* are analogous by virtue of sharing the relational signature *disable-experiencer*, it cannot recursively determine entity mappings like *eye:leg* and *look:walk* in the way that structure-mapping can. The taxonomic approach thus has a uniform mapping richness of 1.

6.2 Evaluating Structure-Mapping

6.2.1 Structure-Mapping Coverage

A structure-mapping signature can be generated for every structured definition in HowNet. In principle then, the coverage of this approach is 100%. In practice, however, 10% of HowNet's semantic definitions contain no real structure beyond the specification of a hypernym or a domain tag. The maximum coverage of structure-mapping then, as limited to definitions with relational structure, is 90%.

6.2.2 Structure-Mapping Recall

HowNet's 21,761 unique structured definitions comprise 21,929 unique propositions. From these, 21,159 unique structural signatures are derived (many of which are generalizations of other signatures), serving to find analogues for 14,370 definitions. The recall of structure-mapping is thus 61%, while the most productive signature is:

{component.部分:whole={?}}

This signature serves to analogically co-index the 397 unique definitions for concepts that exhibit a part-whole distinction.

6.2.3 Structure-Mapping Parsimony/Precision

With 79% of all structural signatures serving to index just a single definition, the parsimony of the structure-mapping approach must be judged as a low 21%. However, the parsimony of the structure-mapping approach does not have the same critical import for HowNet's overall design as does the parsimony of the relational abstraction approach. Here we measure the reusability of structural forms, or patterns, rather than the predicates that semantically anchor these forms. Nonetheless, a higher parsimony score is desirable, and would reflect a higher degree of structural organization in HowNet.

6.2.4 Structure-Mapping Richness

Most analogies (64%) generated using the structure-mapping approach imply two entity mappings, 25% imply three entity mappings, and 11% imply four or more. The average mapping richness of a structure-mapped analogy is thus 2.48.

6.3 Analysis of Results

The results of this comparison, as summarized in Table 1 below, force us to draw some important conclusions about the utility of each approach to performing analogical reasoning in HowNet.

	Abstraction	**Structure-Mapping**	**Combination**
Coverage	.31	.90	.90
Recall	.30	.61	.72
Parsimony	.59	.21	.24
Richness	1.0	2.48	2.24

Table 1: Comparison of both approaches to analogy in HowNet

First, though the abstraction approach is capped by the limited use of self-reference among HowNet definitions, it demonstrates a recall rate that closely approaches this ceiling, managing to find analogies of non-trivial complexity for almost 1 in 3 HowNet definitions (or 1 in 4 without structural inversion). Because of its broader coverage, structure-mapping does considerably better, generating analogies for 3 in 5 definitions. A combination of both approaches ("combination" in Table 1) generates analogies for almost 3 in 4 definitions, which is most encouraging given the creative demands of analogy generation. This is especially so as we have considered here analogies between unique definitions, not unique words. The inherent ambiguity of natural language means that just one inter-definition analogy might be lexically realized in tens, perhaps even hundreds, of different ways.

7 Conclusions

Relational abstraction, structural inversion and structure-mapping are all forms or reconceptualization, since each derives new semantic structures from old. This paper has

explored these three different, but ultimately complementary, approaches to reconceptualization within the specific context of HowNet, a large-scale conceptual resource.

Though reconceptualization involves representational change, it is debatable whether reconceptualization as explored here possesses the radical power of conceptual change ascribed to the process of transformational creativity by Boden (1990, 1999). These approaches to reconceptualization do not dramatically reconfigure the conceptual space in which creative processing is to occur, but the conceptual space is nonetheless modified in an important way. For instance, the first approach, relational abstraction, allows a new taxonomic organization to be constructed from the relational *predicate:case-role:filler* structure of HowNet's propositional semantics. This new taxonomy is based not on conventional categories of being, but on the functional and behavioral nature of the concepts involved. As such, it provides multiple layers of relational abstraction that can facilitate creative analogy between semantically distant, but functionally similar, HowNet entries. This strategy can additionally be seen as a form of *meta*-reconceptualization, since it allows the constructive semantics of HowNet to be construed in the differential manner of WordNet. That is, through the process of relational abstraction, HowNet's conceptual space is altered such that HowNet's semantic structure becomes both constructive and differential. In this light at least, relational abstraction is a transformative process.

The second approach to reconceptualization is based not on abstraction, but on structural rarefaction, though philosophically, both mechanisms are similarly motivated. To the extent that a relational abstraction is based on the combination of a predicate and a case-role, it can be seen as a compressed and very regular form of structural signature. As such, we begin to appreciate that the abstraction and structure-mapping approaches to

analogy are not that different after all. Both aim to reconceptualize a concept in a way that allows important semantic similarities to be highlighted, while unimportant dissimilarities are forced into the background.

The third mechanism of reconceptualization we have explored is structural inversion, which effectively allows a system to look outside a concept to obtain a new semantic perspective from the vantage point of other concepts. The strategy of structural inversion clearly complements that of relational abstraction, since the former provides additional propositional content for the latter to abstract over. In fact, structural inversion often provides multiple alternate perspectives on a concept, any of which might be used to generate an analogy or, more generally, to solve a problem. Consider the concept ***software|软件***, which HowNet simply defines as a kind of ***implement|软件***. Structural inversion allows software to be redefined, among other things, as anything that can compiled via a programming language, or anything that is damaged by a computer virus. This form of redefinition is clearly quite liberal, as not everything affected by a virus is generally deserving of the label "software". Yet, liberal categorization lies at the root of creative thinking: this redefinition forces us to consider web-pages, spreadsheets and even email messages as software, and indeed, under closer examination, all do fit the bill as "soft" wares.

The liberality of structural inversion seems well-suited to the robust treatment of categories whose membership criteria are arbitrary or highly subjective. Consider the concepts treasure, curio and oddity, each of which receive cursory treatment in HowNet's semantics. Structural inversion allows a system to reconceptualize the concept ***treasure|珍宝*** as anything that is stored in a jewellery box, sold in a jewellery shop, or hidden on a treasure

ship. Were someone to store something of subjective value in a jewellery box, such as love letters, photos, etc., a creative system based on structural inversion would certainly be able to recognize their value. Though the relative contribution of structural inversion to the "bottom line" of the HowNet evaluation is relatively slight, enabling the recall of the relational abstraction approach to jump from 26% to 31%, we believe it to be a promising technique that deserves further research in the context of other resources and creative tasks.

In closing, we note that the results of this work, in particular the perspective of relational abstraction, can be tangibly appreciated in the *Analogical Thesaurus*, an on-line index derived from HowNet that allows word-concepts to be retrieved using both analogy and metonymy. This index is available for use on-line at: *Afflatus.ucd.ie*.

References

Attardo, S, Hempelmann, C. F. and Di Maio, S. (2002). Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *Humor: International Journal of Humor Research* 15-1, 3-46.

Boden, M. (1990). *The Creative Mind: Myths and Mechanisms.* London: Weidenfeld and Nicholson.

Boden, M. (1999). Computational models of creativity. *Handbook of Creativity*: 351–373.

Carpuat, M., Ngai, G., Fung, P., and Church, K.W. (2002). Creating a Bilingual Ontology: A Corpus-Based Approach for Aligning WordNet and HowNet. *The Proceedings of GWC 2002, the 1st Global WordNet conference*, Mysore, India.

Dong, Z. (1988). Knowledge Description: What, How and Who? *The Proceedings of*

the International Symposium on Electronic Dictionaries, Tokyo, Japan.

Hutton, J. (1982). *Aristotle's Poetics*. Norton.

Falkenhainer, B., Forbus, K. and Gentner, D**.** (1989). Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41, pages 1-63.

Fass, D. (1988). An Account of Coherence, Semantic Relations, Metonymy, and Lexical Ambiguity Resolution. *In: S. Small, G. W. Cottrell and M. K. Tanenhaus (eds.): Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology and A.I.* Morgan Kauffman: CA.

Kuhn, T. (1962). The *Structure of Scientific Revolutions.* Chicago: University of Chicago Press. J.

McCarthy, J. (1999). Creative Solutions to Problems. In the proceedings of the AISB'99 symposium on AI and Scientific Creativity (Edinburgh Scotland, 1999).

Miller, G. (1995). WordNet: A Lexical Database for English. *Comms. of the ACM,* 38(11).

Ritchie, G. (1999). Developing the Incongruity-Resolution Theory. *In the proceedings of the 1999 AISB Symposium on Creative Language: Stories and Humour,* Edinburgh, Scotland.

Veale, T. and Keane, M.T. (1997). The Competence of Sub-Optimal Structure Mapping on Hard Analogies. *The proceedings of IJCAI'97, the International Joint Conference on Artificial Intelligence,* Nagoya, Japan.

Veale. T. (2003). Dynamic Type Creation in Metaphor Interpretation and Analogical Reasoning: A Case-Study with WordNet. *The proceedings of ICCS2003, the 2003*

International Conference on Conceptual Structures, Dresden, Germany

Veale, T. (2004a). Incongruity in Humor: Root-Cause or Epiphenomenon? *The International Journal of Humor* 17/4, a Festschrift for Victor Raskin

Veale, T. (2004b). WordNet sits the S.A.T.: A Knowledge-Based Approach to Lexical Analogy. *The proceedings of ECAI'2004, the 16th European Conf. on Artificial Intelligence*. John Wiley: London.

Way, E. C. (1991). Knowledge Representation and Metaphor. *Studies in Cognitive systems*. Kluwer Academic Publishers: Holland.

Wong, S.H.S. (2004). Fighting Arbitrariness in WordNet-like Lexical Databases – A Natural Language Motivated Remedy. *The proceedings of GWC 2004, the 2nd Global WordNet conference*, Sojka, Pala, Smrz, Fellbaum, Vossen (eds.).